ATHLETES
UNLIMITED

PARITY IN ATHLETES UNLIMITED SOFTBALL

Parity in Athletes Unlimited Softball

Author: Philip Maymin

Abstract: Athletes Unlimited developed an alternative scoring system for the sport of softball, designed to produce greater team parity and to highlight individual parity. By creating a scoring system that rewards individual contributions as well as team accomplishments, it is believed that players' skill and athleticism is more fairly measured. Simulations also suggest that the rotating drafting order will result in greater team parity. Milestones are calculated for how many non-team points an individual must earn for any given week in order to overcome any deficits from unfortunate team placement.

Introduction

This paper investigates the parity implications of the innovative scoring system to be employed by the Athletes Unlimited (AU) professional softball league.

Traditional scoring systems in similarly scored sports simply count the number of wins. Thus, they do not distinguish between dominant wins in which the same team won every inning and narrow wins in which the teams traded innings. In the latter case, the true parity of the teams is closer than what a traditional scoring system would indicate. To address this issue, the AU team scoring system allocates points to the winner of each inning as well as to the winner of the game overall.

Traditional professional leagues attempt to address team disparity by offering the worst performing team the highest draft pick, or some random-based variation of that. Nevertheless, perhaps because drafting quality and playing quality both relate to some general quality of a team, there continue to exist numerous dynasties across professional sports. The AU system essentially does away with fixed teams by redrafting every roster every week. The four highest ranked individuals are named captains. Thus, absent no other information, presumably team 1, the first captain's team, is likely stronger than the second, and so on down the line. To address the team disparity issue from drafting order, the AU draft proceeds in a rotating style: 1-2-3-4 in the first round, then 2-3-4-1, then 3-4-1-2, and so on.

Finally, ranking individual players in a team game has been a consistently controversial and contentious topic across all sports. The AU individual ranking system comprises three parts: team points, MVP points, and individual points. The team points are as described above. MVP points are awarded to the first, second, and third highest vote receivers among all players in a game. And individual points follow a separate schedule where certain actions result in points; for example, a single is worth 10 points.

Do these innovations work? We address these three primary questions here:

1) What impact does the AU scoring system have on team parity?
2) How does the AU drafting process exacerbate or mitigate team parity?
3) What impact does the AU scoring system have on individual rankings?

Athletes Unlimited Softball

The Athletes Unlimited season lasts 5 weeks, where teams change every week through a player-organized draft. The top 4 athletes on the preceding week's leaderboard are the captains of Team 1 through Team 4. Draft picks are arranged in the following rotation: 1-2-3-4, 2-3-4-1, 3-4-1-2, 4-1-2-3, 1-2-3-4, ... This rotation proceeds every draft round until all team rosters have been filled. Every week, each team plays every other team: thus, there are 6 games in total per week.

The league consists of 56 athletes, with each team comprising 13 athletes in the following positions: 3 pitchers, 2 catchers, 2 middle infielders, 2 corner infielders, 3 outfielders, and one flex/bench player. As a result, 4 of the 56 athletes are not drafted onto a team.

Individual athletes earn points - which translates to bonus compensation - based on the sum of their Win Points, MVP Points, and Individual Points.

Win Points: For every overall game their team wins, athletes earn 50 points each. For every non-overtime inning their team wins, they earn 10 points. Inning points roll over to the subsequent inning in the event of ties, with the exception that overtimes earn no additional points. By construction, a winning team must have earned at least as many Win Points as the losing team.

MVP Points: 60 points are allocated to the player earning the most MVP votes for a given game, 40 points to the second-place player, and 20 points to the third-place player. The players and coaches on both teams of a given matchup, and a small portion of fans, vote on the MVP of the game.

Individual Points: Individual points accumulate as follows based on an athlete's offensive plays in the game:

1) +10 points for walks (including BB, IBB, or HBP)
2) +10 points for sacrifice hits (bunts) or sacrifice flies
3) +10 points for each stolen base (-10 points if caught stealing)
4) +10 points for singles
5) +20 points for doubles
6) +30 points for triples
7) +40 points for home runs
8) Pitchers:
    a. +4 points for each out recorded
    b. -10 points for each earned run allowed

Team Parity

Considering only the Team Points, what impact does the AU scoring system have on team parity?

To answer this question, we simulated a 6-game, 4-team week of softball and compared various scoring systems. We employed the Gini coefficient as our measure of team parity, a commonly used metric of concentration used in evaluating team parity, industry monopolization, and wealth inequality: a Gini coefficient of 1 means all of the value is held by a single entity and the universe is thus highly concentrated, while a Gini coefficient of 0 means no entity holds more than any other entity and the universe is thus highly equal.

Simulations were performed assuming a Poisson process of runs per inning for each team. The Poisson process takes one parameter, corresponding to the average number of runs per inning. This also corresponds to the standard deviation of the number of runs per inning; other distributions generalize to allow for different averages and standard deviations. The Poisson process broadly - but not perfectly - matches empirical distributions for baseball and softball, but the method is conventionally considered a good approximation. A matchup between two teams is the difference between two Poisson distributions, known as the Skellam distribution.

Unequal Talent

Our first scenario allows for the 4 teams to have a difference in their overall "talent", which we defined earlier as our Poisson process parameter:

- Team 1 on average scores 0.7 runs per inning.
- Team 2 on average scores 0.6 runs per inning.
- Team 3 on average scores 0.5 runs per inning.
- Team 4 on average scores 0.4 runs per inning.

These numbers were chosen to be broadly in line with average scores of completed games.

Table 1 below lists these parameters in the Score column. The average rate of scoring across the league is thus 0.55. We can therefore compute the average number of runs allowed by each team. For example, Teams 2, 3, and 4 average 0.5 runs per inning. Thus, Team 1 allows 0.5 runs per inning. Continuing that calculation for each team results in the Allow column. Finally, the Net column is simply the difference for each team between the amount of runs they score on average per inning and the amount of runs they allow on average per inning.

Table 1: Average Team Offense and Defense

Team	Score	Allow	Net
1	0.70	0.50	0.20
2	0.60	0.53	0.07
3	0.50	0.57	-0.07
4	0.40	0.60	-0.20
Average	0.55	0.55	0.00

Table 2 below evaluates the per-inning expectation of the corresponding Skellam distribution for each of the six possible matchups, and also calculates the probability of the first team winning, losing, or tieing that inning.

Table 2: Expected Matchup Outcomes per Inning

Matchup Inning	Expected Score	Win/Lose/Tie Prob.
Team 1 vs Team 2	0.68 - 0.55	34% / 25% / 41%
Team 1 vs Team 3	0.72 - 0.45	38% / 21% / 42%
Team 1 vs Team 4	0.76 - 0.36	42% / 16% / 42%
Team 2 vs Team 3	0.62 - 0.48	32% / 24% / 44%
Team 2 vs Team 4	0.65 - 0.39	36% / 19% / 45%
Team 3 vs Team 4	0.55 - 0.41	31% / 22% / 48%

In the first matchup, Team 1 is playing against the defense of Team 2, which allows 0.53 runs per inning, slightly better than the 0.55 league average. Thus, Team 1 is expected to score a little below its average. Similarly, Team 1 has the league's best defense and holds Team 2 below its scoring average. Even though Team 1 is more talented than Team 2, it only has a 9% higher chance of winning the inning, largely because the single most likely outcome, at 41%, is a tie.

Table 2 above can be extended to an entire game: see Table 3 below. Naturally, the better team has a better chance of winning the entire game than of merely winning a single inning.

Table 3: Expected Matchup Outcomes per Game

Matchup Game	Expected Score	Win/Lose/Tie Prob
Team 1 vs Team 2	4.75 - 3.82	56% / 31% / 13%
Team 1 vs Team 3	5.05 - 3.18	68% / 20% / 12%
Team 1 vs Team 4	5.35 - 2.55	80% / 11% / 9%
Team 2 vs Team 3	4.33 - 3.39	56% / 30% / 14%
Team 2 vs Team 4	4.58 - 2.72	69% / 19% / 12%
Team 3 vs Team 4	3.82 - 2.88	56% / 29% / 15%

We consider these two different types of scoring systems:

1. **Conventional Win/Loss:** the winning team earns one point, the losing team nothing.
2. **AU Scoring:** each team earns 10 points for each inning it outscores the opponent (points rolling over to subsequent innings - except extra innings - in the event of ties), and 50 points for winning the game as a whole.

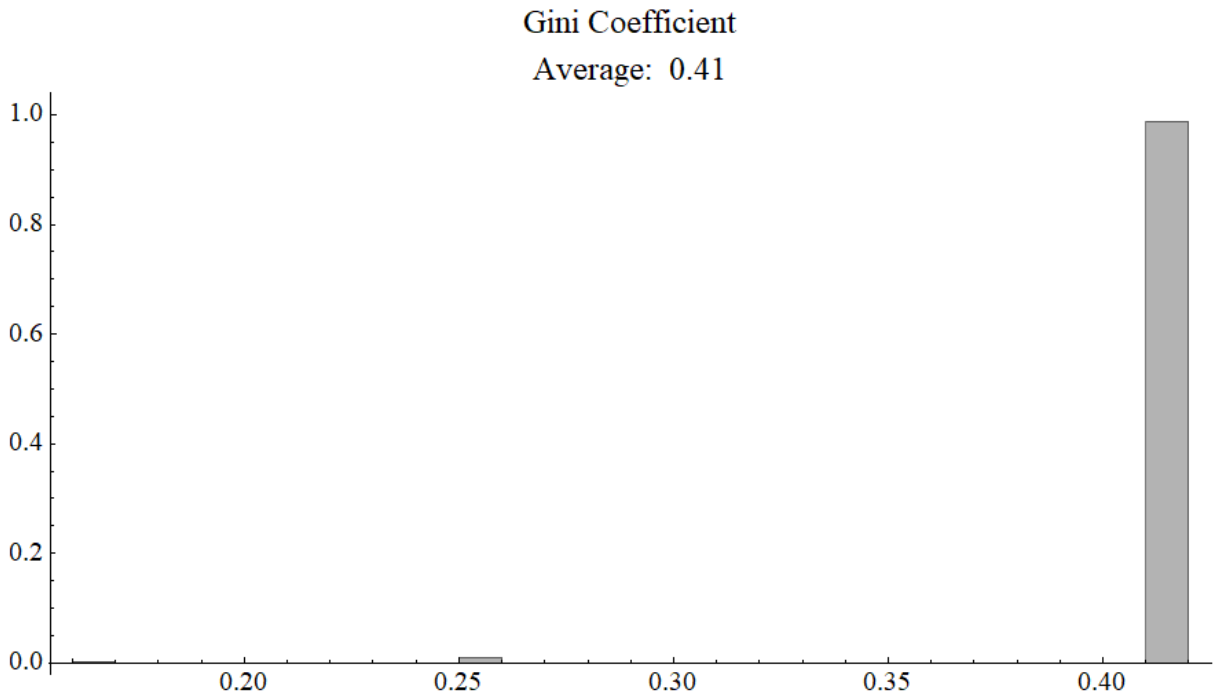
We ran 1,000 simulations per matchup, or 6,000 simulations in total. The teams earn points according to each of the two scoring systems above and we rank the four teams based on the corresponding totals.

Table 4: Conventional Win/Loss Scoring

Rank	First	Second	Third	Fourth
Team 1	84%	16%	0%	0%
Team 2	22%	66%	12%	0%
Team 3	6%	18%	63%	13%
Team 4	0%	12%	13%	75%

The most likely outcome under the conventional win/loss scoring system is that Team 1 ranks first, 84 percent of the time. This results in a very unbalanced league with a single dominant team, even though the team is only marginally better than the next team.

Figure 1: Gini Coefficients for Conventional Win/Loss Scoring Simulations



Each simulation also results in a Gini coefficient. Figure 1 above plots the histogram of these Gini coefficients.

The average Gini coefficient is 0.40, and it is clear that an even more extreme occurrence happens approximately 90 percent of the time, and only the rare weeks when there are more ties or sufficient upsets result in anything resembling team parity.

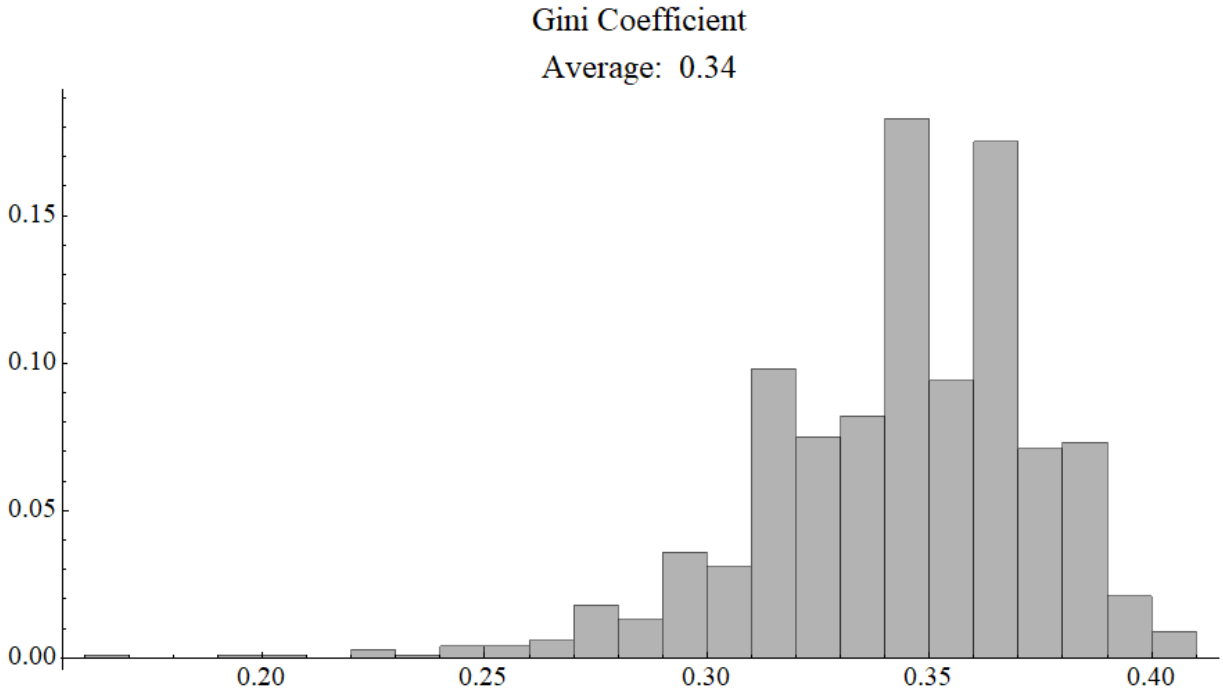
Table 5 and Figure 2 compare the same calculations for the AU scoring system.

Table 5: AU Scoring

Rank	First	Second	Third	Fourth
Team 1	83%	16%	0%	0%
Team 2	17%	69%	14%	0%
Team 3	1%	16%	70%	13%
Team 4	0%	2%	18%	80%

The distribution of ranks are approximately the same, with Team 1 ranking in first place 83 percent of the time. However, because it is far more likely an underdog can win at least some innings even in an eventual loss, the distribution based on Win Points is far more equal.

Figure 2: Gini Coefficients for AU Scoring Simulations



Not only is the Gini coefficient on average substantially lower with the AU scoring system than in the conventional win/loss scoring system, but the distribution is clearly far less discontinuous.

In other words, while the best team still won most often, it won on a different scale: in effect, the best team won by a smaller numerical margin. This is important because it means the unlucky athletes on other teams have an easier chance to catch up to the lucky athletes on the winning teams. The section on individual parity below explores this further.

An interactive website is available to run these simulations with alternate parameters: https://www.wolframcloud.com/obj/philip/au/teamsim

The Draft

Every week, the top 4 players on the AU leaderboard become the captains of the 4 teams and take turns choosing their teammates from the remaining athletes. Which drafting system is best?

To answer this question, we simulated the individual "production" or "softball ability" of each athlete and assumed that all captains always choose the highest ranked player still available.

Professional athletes rank higher than the average human being on some dimension of athletic ability, which we will label “softball ability”. Thus, our athletes are some number of sigmas above the average human being in that metric.

Suppose that this “softball ability” metric is normally distributed and that AU softball players are all at least 5-sigma above the average. Then out of the 250 million females in their early twenties on earth, there are likely to be about 70 individuals who are more than 5-sigma above the mean, of which 56 are in the AU league.

We simulated 56 random "softball ability" metrics that are at least 5 sigma above zero. After sorting them by value, the first 4 are deemed captains. The last 4 are dropped from the draft.

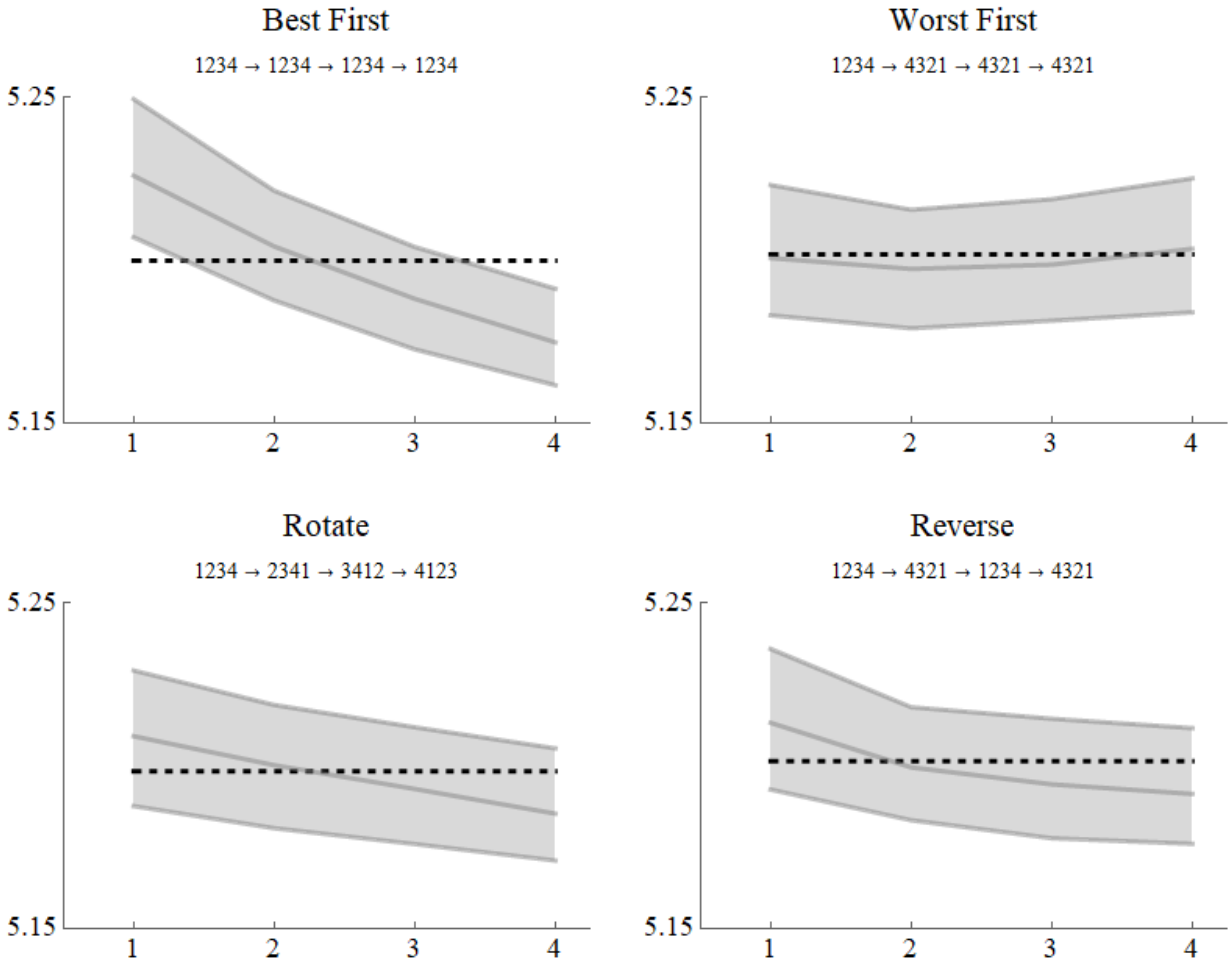
We then identified 4 possible draft pick systems to simulate. In the “Best First” system, team 1 always goes first in every round, followed by team 2, then team 3, and finally team 4. If the best captain always chooses first, their team will tend to be better than team 2, and team 2 better than team 3, and team 3 better than team 4.

If the draft were to take place in the opposite order, with the fourth-best captain choosing first, the resulting parity is much, much better. However, this would potentially create counterintuitive incentives for players to perform worse and attempt to secure the team captain position for team 4 in order to secure better overall draft picks. To minimize this behavior, the first round of the draft should go in the order of the teams: team 1 picks first, team 2 picks second, etc. This is called the “Worst First” system.

The “Reverse” draft pick system attempts to find a medium ground between Best First and Worst First by alternating round starters between team 1 and team 4 and manipulating the pick order accordingly (i.e. 1234 → 4321). Finally, the “Rotate” draft pick system gives every team a first pick per round until the draft is completed (i.e. 1234 → 2341 → 3412 → 4123).

Figure 3 below shows the quartile bands over hundreds of simulations for each draft system.

Figure 3: Team Parity Draft Simulations



The simulations of Figure 3 are for 5-sigma exceptionalism. The results across other sigmas are virtually identical. In all cases, the Worst First (except for round 1) draft strategy has the best overall parity. However, in that system, the fourth team actually does slightly better than the first, in the median. Rotating and Reversing perform about the same: both have mild positive bias to team 1 and mild negative bias to team 4. The maximum deviations from perfect parity are shown on the graph.

In the Best First or Worst First draft systems, each team always waits three picks before picking again. In the Reverse draft system, each team waits either zero, two, four, or six picks before picking again, each with equal probability, one quarter of the time. In the Rotate draft system, each team waits two picks three-quarters of the time and six picks one-quarter of the time, which dominates the Reverse system by effectively collapsing the zero and four pick waits onto the two-pick wait. In other words, for approximately the same level of parity, the Rotate system reduces the variability of the wait times.

Therefore, because it results in team parity that is closest to equal without benefitting later teams more, and because both the amount and the variability of waiting is minimized relative to others, the Rotate draft order appears to be the best overall choice.

Individual Parity

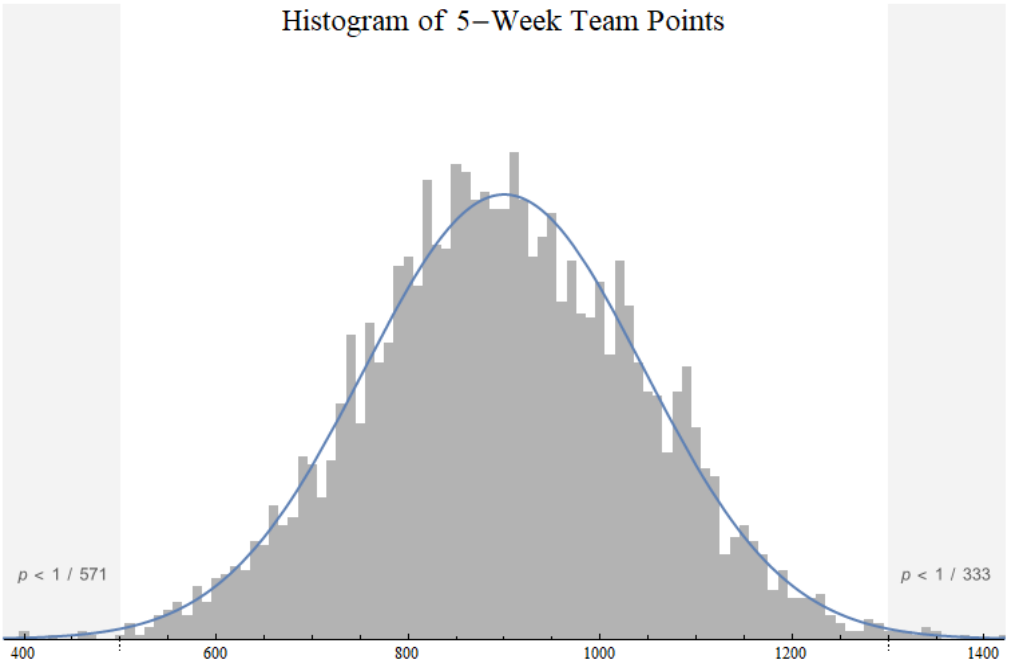
An individual athlete can only control her own production and not the team she is placed on, so one natural concern is how much of a deficit a run of bad draft luck can bring and whether it can be overcome. How likely is an athlete to find herself on the worst team for five straight weeks?

Using the same simulation methodology as for Team Parity above, we simulated five consecutive weeks of games (the equivalent of one AU season) and evaluated the likelihood and impact of being on the worst team each week.

Based on the simulations, there is a less than 1 in 500 chance of consistently being on the worst team, and hence earning merely 500 Win Points. (Similarly, and symmetrically, there is about the same chance of consistently being on the best team and hence earning 1300 Win Points.)

NB: a conventional probability calculation would suggest that there is a $(1/4)^5$ = one in a thousand chance of being on the worst of four teams five straight times, but the simulation accounts for potential ties.

Figure 4: Histogram of Simulated 5-Week Team Points



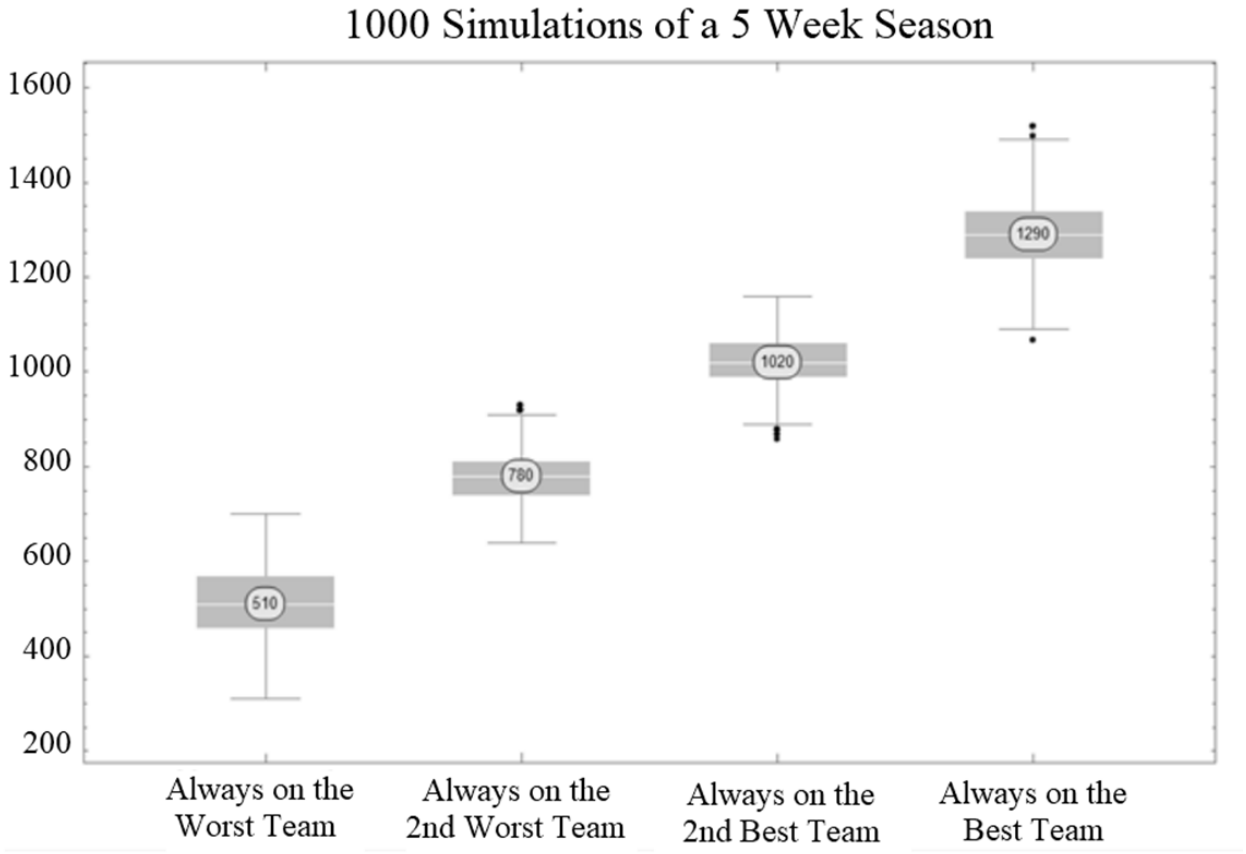
Most players will be within one or two standard deviations (150) of the average (900): the 95% confidence interval is from about 600 to about 1200. That means of the 52 drafted athletes, approximately two or three would be outside that range.

Is that range insurmountable? A difference of 600 points is large, but not impossible. MVP points can overcome the difference. If the best athlete happened by sheer bad luck to find herself on the worst team for five consecutive weeks, but she was still voted MVP for each of her games, she would have gained additional points, placing her well ahead of any non-MVP on the best team.

An alternative way of summarizing the histograms is presented in Figure 5, showing a box-and-whisker chart of the Win Points of individuals who were always on the worst, second worst, second best, or best team.

An athlete who had the misfortune to always be on the worst team the entire season would need about 700 MVP points to eclipse any non-MVPs on the best team all season.

Figure 5: Average, Quartiles, and Outliers of Individuals on Specifically Performing Teams



Breaking it down further, we can look at the results on a cumulative week-by-week basis.

Table 6: Distribution of Individual Win Points

Weeks	Mean	StdDev	2/3 Range		95% Range		MVP1	MVP2	MVP3
1	180	64	116-244	-128	52-308	-257	180	120	60
2	360	91	269-451	-182	178-542	-363	360	240	120
3	540	111	429-651	-222	318-762	-444	540	360	180
4	720	128	592-848	-257	463-977	-513	720	480	240
5	900	145	755-1045	-289	611-1189	-578	900	600	300

Table 6 shows that after 1 week, an MVP on the worst team would NOT be the #1 overall athlete on the leaderboard. But because MVP points grow linearly while randomness only grows with the square root of elapsed time, that changes quickly.

After 2 weeks, the MVP of the always-worst team would be approximately tied with an average player on the always-best team. After 3 weeks, the MVP of the always-worst team would be the leader. After 4 weeks, she would be far and away the leader, and after 5 weeks, even someone who continuously got second place MVP votes, rather than first place, would still be the overall leader, compared to people with no MVP votes.

These findings indicate that a player is not necessarily destined for the bottom of the leaderboard if she finds herself on the losing team, or even on a string of losing teams. Individual points - particularly MVP points - allow for players to be acknowledged and rewarded for exceptional athletic performance, even if their teams ultimately fail to perform at the same level.

Conclusion

We have used simulations to address team parity in the new AU scoring system. Specifically, we have addressed the following three questions:

1) What impact does the AU scoring system have on team parity?

The AU scoring system results in greater team parity than traditional win/loss scoring methods. Not only does it result in greater team parity on average, but the variability of the team parity is reduced as well.

2) How does the AU drafting process exacerbate or mitigate team parity?

The AU "rotating" drafting system further mitigates team parity relative to the more traditional fixed-order format. However, it does not reach total team parity, otherwise there would be no advantage to finishing in first place. Indeed, compared to other possible drafting systems, the AU system incentives a higher draft order for each of the four teams. In other words, under the AU system, it is never to a potential captain's advantage to secure a lower draft order.

3) What impact does the AU scoring system have on individual rankings?

Individual rankings can differ depending on the good or bad "luck" of consistently being on a winning or losing team for several weeks in a row. However, the difference is easily bridged with MVP votes, before even taking into account individual points. In other words, if one player is clearly the best in the league but happens to end up on the worst team several weeks in a row, she will still earn enough MVP points to be at the top of the leaderboard.

About the Author

Dr. Philip Z. Maymin is a professor of analytics and the director of the Master of Science in Business Analytics program at the Fairfield University Dolan School of Business. He is the founding managing editor of Algorithmic Finance and the co-founder and co-editor-in-chief of the Journal of Sports Analytics. He is the Chief Technology Officer for the Esports Development League (ESDL), an Insight Partner with Essentia Analytics, an advisor to Athletes Unlimited, and an affiliate of the Langer Mindfulness Institute, and has been an analytics consultant with several NBA teams.

He holds a Ph.D. in Finance from the University of Chicago (dissertation chair: Richard H. Thaler), a Master's in Applied Mathematics from Harvard University, and a Bachelor's in Computer Science from Harvard University. He also holds a J.D. and is an attorney-at-law admitted to practice in California.

He has been a portfolio manager at Long-Term Capital Management, Ellington Management Group, and his own hedge fund, Maymin Capital Management.

He has also been a policy scholar for a free market think tank, a Justice of the Peace, a Congressional candidate, a professor of finance and risk engineering at NYU, a professor of analytics and finance at the University of Bridgeport, and an award-winning journalist and columnist.

He was a finalist for the 2010 Bastiat Prize for Online Journalism. He was awarded a Wolfram Innovator Award in 2015. He won the Wolfram Live Coding Challenge in 2016 and second place in 2018, and he won the Wolfram One-Liner Competition in 2015, 2016, 2018, and 2019. He was named one of the Top 50 Data and Analytics Professionals in the US and Canada by Corinium in 2018. He is the only person to have won both the Grand Prize for Best Research Paper (2018) and the Hackathon (2020) at the MIT Sloan Sports Analytics Conference.

His popular writings have been published in dozens of media outlets ranging from Bloomberg to Forbes to the New York Post to American Banker to regional newspapers, and his research has been profiled in dozens more, including The New York Times, Wall Street Journal, USA Today, Financial Times, Boston Globe, NPR, BBC, Guardian (UK), CNBC, Newsweek Poland, Financial Times Deutschland, and others.

His research on behavioral and algorithmic finance has appeared in Quantitative Finance, North American Journal of Economics and Finance, Journal of Portfolio Management, Journal of Wealth Management, Journal of Applied Finance, and Financial Markets and Portfolio Management, among others, and his textbook Financial Hacking was recently published by World Scientific.